Chaos, Solitons and Fractals 96 (2017) 43–50

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

A novel method of evaluating key nodes in complex networks

Canshi Zhu[a], Xiaoyang Wang[b,*], Lin Zhu[c]

[a] Information Engineering College, Xijing University, Xi'an 710123, China
[b] Xi'an Satellite Control Center, Xi'an 710043, China
[c] Equipment Management and Safety Engineering College, Air Force Engineering University, Xi'an 710051, China

ARTICLE INFO

Article history:
Received 3 March 2016
Revised 13 January 2017
Accepted 14 January 2017

Keyword:
Complex networks
Structural hole
Closeness centrality
Influence matrix

ABSTRACT

In order to evaluate the influence of nodes in complex networks, a new method is advanced of evaluating key nodes in complex networks, in combination with the "structural hole" theory and closeness centrality of nodes, through defining and applying the influence matrix of nodes' "structural holes" in response to the limitations of existing methods. The "structural hole" theory gives a comprehensive consideration of the node degree as well as information about topological relations with its neighbors, whereas the closeness centrality of nodes is a reflection of the node's global information. The "structural holes" influence matrix in degree reflects the node's local and global information. So a more proper evaluation standard is established for influence of nodes and a simulation analysis is made of different-scale networks. The results of such analyses show that the method can not only make an exact assessment of the influence of nodes, but also obtain ideal evaluation results from actual complex networks of different scale.

1. Introduction

Node influence evaluation in complex networks is a fundamental problem in network analysis, and it has become a hot subject in the present research [1,2]. So it is very important to determine the key nodes in different networks, such as the power network [3], the traffic network [4,5], the process of controlling viruses and disease immunities [6,7] and so on.

However, evaluation of key nodes in networks of different scale requires different approaches. At present, the node influence evaluation in a network structure is based mainly on four aspects [8]—local property, global property, location and random walk of the network. Among them, the local property of the network is mainly concerned with the information of the node in its own as well as of its neighbors, and, since it is simple in calculation and low in time complexity, it is usually applicable to large-scale networks. According to Reference [9], for example, the influence of nodes in the network bears only a certain relation to the extent of their own. It considers that the node influence comprises its initial influence and the influence contributions from both the adjacent and nonadjacent nodes according to the dependence strength between them. Reference [10] researches the spreading ability of nodes in complex networks based on local structure. It proposed a local structural centrality measure which considers both the number and the topological connections of the neighbors of a node. Reference [11] gives a comprehensive consideration of the number of the node's neighbors as well as the close degree of their connection; and thus, a method is proposed of assessing the influence of the node based on neighbor information and clustering coefficient. The global property of the network is mainly concerned with the global information of the network, but high in time complexity, it is not applicable to the large-scale network. According to Reference [12] the betweeness centrality can be a very good measure of a node's influence. Indeed more a node is characterized by a shorter geodesic distance from all the other nodes of a network, higher it would be its influence on the network. Reference [13] introduces the concept of shortest path and betweenness centrality, and thinks the betweenness centrality approach takes into account of both the fluxes between species and their relative positioning within the chemical network. Kitsak et al [14] put forward, for the first time in 2010, the idea that influence of nodes is dependent on their position in the whole network, and, by means of the k-shell decomposition, obtained the ranking index for influence of nodes. The index, low in time complexity, is applicable to large-scale networks, and, due to degree and betweenness, is better able to identify the most influential nodes in the spread of the disease. Reference [15] thinks the k-shell index, which is the topological location of a node in a network, is a more efficient measure at capturing the spreading ability of a node than are the degree and betweenness centralities. And it proposed coreness centrality, to estimate the spreading influence of a node in a network using the k-shell indices of its neighbors. Reference [16] puts forward an indicator on the basis of k-shell decomposition and the Technique

* Corresponding author.
E-mail address: wangxiaoyang1987@163.com (X. Wang).

http://dx.doi.org/10.1016/j.chaos.2017.01.007
0960-0779/

for Order Preference by Similarity to Ideal Object, to evaluate the node influence comprehensively. The method of ranking the node influence of random walk is mainly based on the ranking technology used for linking the relationship between web pages. Just as the linking relationship between web pages can be explained as correlation and mutual support between web pages, so also the influence of nodes can be identified. In Reference [17], the PageRank algorithm is adopted to establish the classification of nodes, and in combination with the concept of betweenness centrality in random walk, a method of central measurement is advanced to sort the urban network nodes. Such typical methods also include PageRank [18,19] algorithm, LeaderRank [20,21] algorithm, HITS [22] (hypertext induced topic search) algorithm, etc..

The methods, as mentioned above, evaluate node influence or sort nodes only from a certain aspect. As a matter of fact, the influence of network nodes is closely related not only to the local property of nodes, but also to the location of nodes in the network as well as to dependence on each other [23]. Reference [24] also pointed out that the calculation based on single index in different network topologies is more liable to be one-sided. The influence of a node in the network is related to the overall structure of the network, so it is necessary to make a comprehensive evaluation of its influence through more than one influential index. Thus, evaluation of a node's influence should not only consider the node's own attributes, but also consider the global properties of the node. Thereby, it is possible to evaluate key nodes of the networks in more accurate and efficient manners combining local information and global information. In Reference [25], based on the node's local influence (degree), the concept of influence contribution matrix is put forward, complete with information of its location (betweenness), so as to represent interdependence between nodes and serve to evaluate node influence. However, the node degree can not reflect the topological relationship between the concerned node and its neighbors. In fact, the most influential nodes in the network with community structure should have the following characteristics: the node that plays the role of "bridging" community centers and diverse communities. The calculation of network constrain coefficient embodies the degree attribute and bridging attribute. So, network constrain coefficient embodies the characteristics we are interested to. Therefore, Reference [26] came up with a measuring method based on the local centrality of the structural holes among the node and its neighborhood. The method takes comprehensive account of the number of neighboring nodes as well as of the topological structure between them. Reference [27] also believes that the ranking of key nodes should neither focus on the core nodes in the network nor ignore the nodes in the position of the structural hole. Reference [28] calculate four centrality measures (closeness, betweenness, degree and PageRank) for authors in coauthorship network and find that centrality measures can be useful indicators for impact analysis. Reference [29] pointed out that the closeness centrality of the node can better reflect the influence of the node on other nodes, and it also reflects the difference of the node's location in the network topology.

To sum up, this paper presents a method to evaluate the key nodes of the complex network through influence matrix of structural holes. This method gives a comprehensive consideration of the node's local information (constraint coefficient) and global information (closeness centrality), thereby making it possible to evaluate key nodes of the networks in more accurate and efficient manners through constructing the "structural holes" influence matrix. In this paper the standard evaluation of key nodes is interpreted as their broadest influence in the network. In terms of communication spreading, their influence corresponds to the maximum scope of information dissemination they can achieve in the whole network. Then this algorithm is compared respectively with degree centrality, betweenness centrality and N-Burt [26] algorithms in authentication network [26], ARPA network and Karate network. And finally, a simulation test is conducted on the spread process in the SIR model. Compared with other algorithms in large-scale networks, this algorithm turns out to be efficient and reliable.

2. Algorithm for evaluation of key nodes

2.1. Theoretical basis

Suppose that the undirected and unweighted network is expressed in Figure $G = (V, L)$, wherein, Figure G contains, n nodes, m edges, $V = \{v_1, v_2, \ldots, v_n\}$ stands for the set of all nodes in the network, $L = \{L_1, L_2, \ldots, L_m\} \subseteq V \times V$ stands for the set of edges in the network. Based on the set L, adjacency matrix $A = [a_{ij}]_{n\times n}$ of the network G can be built, where:

$$a_{ij} = \begin{cases} 1, & i \text{ and } j \text{ have connection} \\ 0, & i \text{ and } j \text{ no connection} \end{cases} \tag{1}$$

Definition 1. Closeness centrality

Closeness centrality can be expressed as the reciprocal of the mean of the distance sum between node i and all the other nodes. Closeness centrality serves to measure the effect of influence the node in the network exerts on other nodes through the medium of network. The greater the closeness centrality of the node is, the greater extent to which the node settles down in the center of the network, and the more influential the node is. Closeness centrality is concretely defined as follows [30]:

$$C_c(i) = (N-1) / \sum_{j=1}^{N} d_{ij} \tag{2}$$

Wherein, N is the number of all nodes, and d_{ij} stands for the shortest distance between node i and node j. Closeness centrality is dependent on the topological structure of the network, and able to precisely detect the central node in response to networks with a star-like structure.

Definition 2. Structural holes

If there exists neither a direct connection between two individuals or between two groups nor an indirect redundancy relationship between them, the hinders between them are known as structural holes. Burt put forward the network constraint coefficient for calculation of structural holes in order to measure the closeness and the structural holes of network [31].

$$p_{ij} = a_{ij} / \sum_{j\in\Gamma(i)} a_{ij} \tag{3}$$

Wherein, p_{ij} represents the proportion of the total effort the node i makes to maintain its neighborhood with the node j, and $\Gamma(i)$ represents the set of the nodes adjacent to the node i. The calculation for the network constraint coefficient of the node i is expressed in:

$$C_i = \sum_{j\in\Gamma(i)} \Big(p_{ij} + \sum_q p_{iq} p_{qj}\Big)^2 \quad q \neq i, j \tag{4}$$

Wherein, q stands for the indirect node that links the node i with the node j, p_{iq} and p_{qj} stand respectively for the proportion of the total effort the node i and the node j make to maintain their neighborhood with their common neighbor q. The smaller the network constraint coefficient is, the greater the structural hole is, and the greater influence the node produces in the spread of information, and the more influential its location is.